

UTILITY PATENT APPLICATION TRANSMITTAL
(Large Entity)*(Only for new nonprovisional applications under 37 CFR 1.53(b))*Docket No.
POU9-2000-0017-USTotal Pages in this Submission
72 (Exc. references)**TO THE ASSISTANT COMMISSIONER FOR PATENTS**Box Patent Application
Washington, D.C. 20231

Transmitted herewith for filing under 35 U.S.C. 111(a) and 37 C.F.R. 1.53(b) is a new utility patent application for an invention entitled:

**TOPOLOGY PROPAGATION IN A DISTRIBUTED COMPUTING ENVIRONMENT WITH NO TOPOLOGY
MESSAGE TRAFFIC IN STEADY STATE**

and invented by:

Knop et al.If a **CONTINUATION APPLICATION**, check appropriate box and supply the requisite information:☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: _____

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: _____

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: _____

Enclosed are:

Application Elements

1. ☒ Filing fee as calculated and transmitted as described below
2. ☒ Specification having 35 pages and including the following:
 - a. ☒ Descriptive Title of the Invention
 - b. ☐ Cross References to Related Applications *(if applicable)*
 - c. ☐ Statement Regarding Federally-sponsored Research/Development *(if applicable)*
 - d. ☐ Reference to Microfiche Appendix *(if applicable)*
 - e. ☒ Background of the Invention
 - f. ☒ Brief Summary of the Invention
 - g. ☒ Brief Description of the Drawings *(if drawings filed)*
 - h. ☒ Detailed Description
 - i. ☒ Claim(s) as Classified Below
 - j. ☒ Abstract of the Disclosure

UTILITY PATENT APPLICATION TRANSMITTAL
(Large Entity)

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.
POU9-2000-0017-US1

Total Pages in this Submission
72 (Exc. references)

Application Elements (Continued)

3. ☒ Drawing(s) *(when necessary as prescribed by 35 USC 113)*
- a. ☒ Formal Number of Sheets Seventeen (17)
- b. ☐ Informal Number of Sheets _____
4. ☒ Oath or Declaration 3 Declarations
- a. ☒ Newly executed *(original or copy)* ☐ Unexecuted
- b. ☐ Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional application only)*
- c. ☒ With Power of Attorney ☐ Without Power of Attorney
- d. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application,
see 37 C.F.R. 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference *(usable if Box 4b is checked)*
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied
under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby
incorporated by reference therein.
6. ☐ Computer Program in Microfiche *(Appendix)*
7. ☐ Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all must be included)*
- a. ☐ Paper Copy
- b. ☐ Computer Readable Copy *(identical to computer copy)*
- c. ☐ Statement Verifying Identical Paper and Computer Readable Copy

Accompanying Application Parts

8. ☐ Assignment Papers *(cover sheet & document(s))*
9. ☐ 37 CFR 3.73(B) Statement *(when there is an assignee)*
10. ☐ English Translation Document *(if applicable)*
11. ☒ Information Disclosure Statement/PTO-1449 ☒ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Acknowledgment postcard
14. ☒ Certificate of Mailing
- ☐ First Class ☒ Express Mail *(Specify Label No.):* EK711552689US

UTILITY PATENT APPLICATION TRANSMITTAL
(Large Entity)

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.
POU9-2000-0017-US1

Total Pages in this Submission
72 (Exc. references)

Accompanying Application Parts (Continued)

15. ☐ Certified Copy of Priority Document(s) (if foreign priority is claimed)

16. ☐ Additional Enclosures (please identify below):

Fee Calculation and Transmittal

CLAIMS AS FILED

For	#Filed	#Allowed	#Extra	Rate	Fee
Total Claims	36	- 20 =	16	x \$18.00	\$288.00
Indep. Claims	3	- 3 =	0	x \$78.00	\$0.00
Multiple Dependent Claims (check if applicable) <input type="checkbox"/>					\$0.00
BASIC FEE					\$690.00
OTHER FEE (specify purpose)					\$0.00
TOTAL FILING FEE					\$978.00

- ☐ A check in the amount of _____ to cover the filing fee is enclosed.
- ☒ The Commissioner is hereby authorized to charge and credit Deposit Account No. **09-0463 (IBM)** as described below. A duplicate copy of this sheet is enclosed.
- ☒ Charge the amount of **\$978.00** as filing fee.
- ☒ Credit any overpayment.
- ☒ Charge any additional filing fees required under 37 C.F.R. 1.16 and 1.17.
- ☐ Charge the issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance, pursuant to 37 C.F.R. 1.311(b).


Signature

Lawrence D. Cutter, Esq.
Reg. No. 28,501
IBM Corporation
Intellectual Property Law
2455 South Rd., P386
Poughkeepsie, NY 12601-5400
Telephone: (914)433-1172
Facsimile: (914)432-9601

Dated: May 30, 2000

CC:

Docket Number: POU9-2000-0017-US1

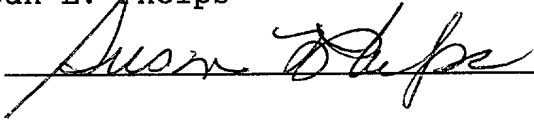
TOPOLOGY PROPAGATION IN A
DISTRIBUTED COMPUTING
ENVIRONMENT WITH NO TOPOLOGY
MESSAGE TRAFFIC IN STEADY
STATE

APPLICATION FOR UNITED STATES
LETTERS PATENT

"Express Mail" Mailing Label No.: EK711552689US
Date of Deposit: May 30, 2000

I hereby certify that this paper is being deposited with the United States Postal Service as "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Box Patent Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Name: Susan L. Phelps

Signature: 

INTERNATIONAL BUSINESS MACHINES CORPORATION

TOPOLOGY PROPAGATION IN A DISTRIBUTED
COMPUTING ENVIRONMENT WITH NO TOPOLOGY
MESSAGE TRAFFIC IN STEADY STATE

Technical Field

5 The present invention relates in general to
communications networks, and more particularly, to a
technique for maintaining a common network topology database
at different nodes in a distributed computing environment
wherein the topology propagation facility generates no
10 message traffic when the distributed computing environment
is in steady state.

Background of the Invention

 A communications network can be generally defined as a
collection of network nodes and end nodes interconnected
15 through communications links or transmission groups. A
network node can be characterized as a data processing
system that provides certain functions within the network,
such as routing of messages between itself and its adjacent
or neighboring nodes, selection of routes for messages to be
20 transmitted between a network node and an end node and the
furnishing of directory services to connected end nodes.
The links between nodes may be permanent communications
links such as conventional cable connections or links that
are enabled only when needed, such as dial-up telephone
25 connections. End nodes are exemplified by devices such as
display terminals, intelligent workstations, printers and
the like which do not provide routing or route selection or
directory services to other nodes in the network.
Collectively, the network nodes, the end nodes and the

transmission groups between the nodes are referred to as network resources. The physical configuration and characteristics of the various nodes and links (and their state) in a network are said to be the topology of the
5 network.

Before a message can be transmitted between any two nodes in any network, a human operator or data processing equipment responsible for establishing the connections needs an accurate and up-to-date file or database on the network
10 topology.

Successful attempts have been made to have the network equipment itself take over the task of maintaining a topology database without human intervention. For example, each processor performing a communication control function
15 can notify other processors of changes in the status of its resources. The other processors use these topology update messages to amend or change their own copies of the topology network database.

In a distributed computing system, several networks may
20 connect the nodes that comprise the system. It is possible that not all nodes are connected to all networks, and multiple "hops" may be needed to transmit messages between any two nodes that are not connected to the same network. To accomplish this, all nodes within the system must know
25 the current global network topology. The topology information includes the set of nodes and network adapters that are connected to each of the networks in the system, as well as the set of adapters and networks that are down. The topology information changes each time a node, network, or

network adapter fails or is powered up. Using the global network topology, each node is able to compute the set of reachable nodes and the route to each reachable node.

5 A need exists in the art for an enhanced technique for disseminating the global topology information to all nodes in the system. More particularly, there is a need for an enhanced topology propagation technique which ensures that there is no propagation of topology messages within the distributed computing environment when the system is in
10 steady state, that is, when no nodes or network adapters fail or are powered up. Preferably, this enhanced technique is achieved without the use of explicit message acknowledgments. The present invention is directed to providing such a topology propagation mechanism.

15 Disclosure of the Invention

To briefly summarize, the present invention comprises in one aspect a method of topology propagation in a distributed computing environment. The method includes:
20 sending group connectivity messages from at least one group leader to identified nodes of at least one group of nodes within the distributed computing environment; discontinuing the sending of group connectivity messages during a time period of no topology change within the distributed computing environment; and reinitiating sending of group
25 connectivity messages from the at least one group leader upon identification of a topology change within the distributed computing environment.

In another aspect, the method includes a system for topology propagation in a distributed computing environment. The system includes means for sending group connectivity messages from at least one group leader to identified nodes
5 of at least one group of nodes within the distributed computing environment, and means for discontinuing the sending of group connectivity messages during a time period of no topology change within the distributed computing environment. The system further includes means for
10 reinitiating sending of group connectivity messages from the at least one group leader upon identification by the at least one group leader of a topology change within the distributed computing environment.

In a further aspect, the invention includes at least
15 one program storage device readable by a machine, tangibly embodying at least one program of instructions executable by the machine to perform a method of topology propagation in the distributed computing environment. The method includes: sending group connectivity messages from at least one group
20 leader to identified nodes of at least one group of nodes within the distributed computing environment; discontinuing the sending of group connectivity messages during a time period of no topology change within the distributed computing environment; and reinitiating sending of group
25 connectivity messages from the at least one group leader upon identification of a topology change within the distributed computing environment.

To restate, provided herein is a topology propagation facility which generates no message traffic when the
30 distributed computing environment employing the facility is

in steady state. The environment is in steady state when there are no failing nodes, network adapters, or network connections, or there are no nodes, network adapters or network connections currently being activated. The topology propagation facility is achieved without the use of explicit message acknowledgments to transmission of topology messages. The topology propagation approach presented herein works in a distributed computing environment comprising multiple networks and multiple adapters, as opposed to existing propagation techniques which assume two-node links. Further, the approach presented herein works well with unreliable networks, i.e., work well without the need for end-to-end acknowledgments. Also, the method presented provides automatic transmission of network topology to a node that is starting up within a distributed computing environment.

Additional features and advantages are realized through the techniques of the present invention. Other embodiments and aspects of the invention are described in detail herein and are considered part of the claimed invention.

Brief Description of the Drawings

The above-described objects, advantages and features of the present invention, as well as others, will be more readily understood from the following detailed description of certain preferred embodiments of the invention, when considered in conjunction with the accompanying drawings in which:

FIG. 1 is a diagram of a representative communications network to employ a topology propagation facility in accordance with the present invention;

FIGS. 2A-2F depict one embodiment of a JOIN protocol employed by multiple nodes of a distributed processing system, wherein FIG. 2A depicts a PROCLAIM message, FIG. 2B a JOIN message, FIG. 2C a prepare to commit (PTC) message, FIG. 2D a prepare to commit acknowledgment (PTC_ACK) message, FIG. 2E a commit broadcast (COMMIT_BCAST) message, and FIG. 2F a COMMIT message and a commit broadcast acknowledgment (COMMIT_BCAST_ACK) message;

FIG. 2G depicts a new group of nodes formed after completion of the JOIN protocol of FIGS. 2A-2F;

FIGS. 3A-3C depict one embodiment of a DEATH protocol employed by multiple nodes of a distributed processing system, wherein FIG. 3A depicts an initial state of the group showing a heartbeat ring, FIG. 3B depicts sensing unresponsiveness of a node and transmitting of a DEATH message responsive thereto, and FIG. 3C depicts a new prepare to commit (PTC) message being sent from the group leader (GL) to the surviving nodes of the group;

FIGS. 4A-4C depict one embodiment of node reachability protocol for a distributed computing environment comprising two networks of nodes, wherein FIG. 4A depicts transmission of a NODE_CONNECTIVITY message to the group leader of network 1, FIG. 4B depicts transmission of a GROUP_CONNECTIVITY message from the group leader to the nodes of the group, and FIG. 4C depicts forwarding of the

GROUP_CONNECTIVITY message from node 2 through its adapter to nodes 4 & 5 of network 2 of the distributed computing environment;

FIG. 5A depicts an initial distributed computing environment to employ a message propagation facility in accordance with the principles of the present invention;

FIG. 5B depicts the initial network connectivity table (NCT) at node 5 of the distributed computing environment of FIG. 5A;

FIG. 5C depicts node 2 disappearing from the adapter membership group (AMG) of FIG. 5A, resulting in a new membership group AMG A_2;

FIG. 5D depicts the NCT at node 5, which is connected to network 2, commensurate with disappearing of node 2;

FIG. 5E depicts the distributed computing environment of FIG. 5C showing propagation of a group connectivity message (GCM) from the group leader of the new adapter membership group A_2 to the active members thereof, and the forwarding of this GCM by node 3 to nodes 5 & 6 of AMG B_1 on network 2;

FIG. 5F depicts the NCT at node 5 upon receipt of the GCM forwarded by node 3 in FIG. 5E;

FIG. 6A is a diagram of one embodiment of a distributed computing environment to employ topology propagation in

accordance with the principles of the present invention,
wherein node 2 is awaiting startup;

FIG. 6B depicts the NCT at node 5 and node 2 pending
startup of node 2;

5 FIG. 6C depicts the distributed computing environment
of FIG. 6A after startup of node 2 and forming of new AMG
A_2;

FIG. 6D depicts the NCT at node 5 and the NCT at node 2
commensurate with startup of node 2, but before updating of
10 the NCTs at the nodes;

FIG. 6E depicts the distributed computing environment
of FIG. 6C showing transmission of a GCM by the group leader
of AMG A_2 and the forwarding thereof by node 3 to nodes 5 &
6;

15 FIG. 6F depicts the NCT at node 5 and the NCT at node 2
following transmission of the GCM from GL1;

FIG. 6G depicts the distributed computing environment
of FIG. 6E showing transmission of a GCM from GL2 (i.e.,
node 6) and the forwarding thereof to nodes 2 & 1 by node 3;

20 FIG. 6H depicts the NCT at node 5 and the NCT at node 2
following publication of the GCM by GL2; and

FIG. 7 is a state diagram for a group leader
implementing a topology propagation facility in accordance
with the principles of the present invention.

Best Mode for Carrying Out the Invention

Generally stated, provided herein is a facility to disseminate global topology information to all nodes in a distributed computing environment, including a heterogenous environment comprising multiple communications networks. This mechanism allows each node to obtain a global view of the system topology, including which network adapters are down, and which nodes belong to partitioned networks. When the global topology stops changing, all nodes in the system will shortly have a mutually consistent view of the topology. Advantageously, no topology propagation messages are sent when the system is in steady state, i.e., when no nodes or network adapters are failing or being added to the environment. Further, topology propagation is achieved without the use of explicit topology message acknowledgments.

As used herein, each network forms an adapter membership group (AMG) with a node chosen as a group leader (GL). GLs and group members periodically send to each other topology propagation messages. These propagation messages are forwarded from a network to other networks if needed. Stopping criteria are applied so that no messages are exchanged when there are no changes in the distributed computing environment. Resuming criteria are also presented to resume topology propagation messages when there are changes in the distributed environment.

A topology propagation facility is described herein in the context of International Business Machines' "Reliable Scalable Cluster Technology" (RSCT) Topology Services

(reference "RS/6000 High Availability Infrastructure," IBM Publication No. SG24-4838-00 ("Redbook") 1996), which is a subsystem responsible for monitoring health of nodes and network adapters in a distributed computing system. This
5 subsystem exists in the IBM RS/6000 SP System or a network of RS/6000 machines. The subsystem is used as a foundation for distributed applications that need to react to failing nodes and other changes in network topology.

FIG. 1 depicts one example of a distributed computing
10 environment, generally denoted 10, showing physical connectivity between multiple nodes 12 across multiple networks 14. Each node 12 is connected to a different network 14 across a unique network adapter 16. As shown, different nodes 12 can have different numbers of network
15 adapters and be connected to different networks 14. In distributed computing environment 10, each node 12 can talk to each other node either directly across a shared network or by "hopping" from one network to another across a node that is common to both networks. The distributed computing
20 environment 10 is assumed to comprise an unreliable computing environment in that messages may be lost in transmission.

As noted, a node may have multiple adapters, each connected to a different network. (Networks may comprise
25 one or multiple sub-nets; and networks may or may not be connected to other networks.) Each adapter in a node, if "up", will be part of a different "adapter membership group (AMG)", since each AMG comprises all the "up" adapters in a network that can communicate with each other.

Logically, it is the "adapter" that is a group leader (GL) of a network: it is possible for a node to have an adapter which is the group leader in one network and another adapter which is not the group leader in its network. On
5 the other hand, it is the node, not the adapter, which runs the topology services daemon, where the protocols are implemented. The daemon implements the protocols on a per-adapter basis.

To simplify the presentation, in FIGS. 2A-3C, the
10 adapter membership protocols are explained in a single-network setting. In this setting, it is simpler to consider the "node" as the group leader. In a multiple network environment, however, it is more precise to consider the "adapter" as the group leader. Depending upon the usage,
15 the term "node" should be read to include its "adapter" when the node is referred to as a group leader.

In one embodiment, each node has a topology services "daemon" process running. This process handles certain aspects of the topology propagation facility of the present
20 invention, including: (1) sending and receiving protocol messages; and (2) storing the global network topology and information about connectivity to the networks to which all the node's network adapters are connected.

In order to monitor the health and connectivity of the
25 adapters in each network, all adapters in the network should attempt to form an "adapter membership group" (AMG), which is a group containing all network adapters that can communicate with each other in the network.

Note that each node may belong to several AMGs, one for each of its network adapters.

To determine the set of adapters that are alive in each network, an adapter membership protocol is run in each of
5 the networks.

As explained further below, adapters that are alive form an Adapter Membership Group (AMG), where members are organized (by way of example only) in a virtual ring topology. To ensure that all group members are alive, each
10 periodically sends "HEART BEAT" messages to its "downstream neighbor" and monitors "HEART BEAT" messages from its "upstream neighbor". Protocols are run when adapters fail or when new adapters become functional. The goal of such protocols is to guarantee that the membership group contains
15 at each moment all the adapters that can communicate with each other.

Each group has a "Group Leader" (GL) and a "Backup Group Leader." The group leader is responsible for coordinating the group protocols, and the backup group
20 leader is responsible for taking over the group leadership when the group leader adapter fails. Both the choice of group leader and backup group leader and the position of the adapters in the ring are determined by a predefined adapter priority rule, which can be chosen to be the adapters' IP
25 addresses. For example, a higher IP address indicates a higher priority.

A list of all possible adapters in each network is contained in a configuration file that is read by all the nodes at startup and at reconfiguration time.

Referring to FIGS. 2A-2G, in order to attract new
5 members to the group, the group leader in each group periodically sends "PROCLAIM" messages to adapters that are in the adapter configuration but do not belong to the group (see FIG. 2A). These messages are only sent to adapters having a lower IP address than that of the sender.

10 The "PROCLAIM" messages are ignored by all adapters that are not group leaders. As shown in FIG. 2B, a group leader node receiving a "PROCLAIM" message from a higher priority (higher IP address) node responds with a "JOIN" message on behalf of its group. The message contains the
15 membership list of the "joining group".

A node receiving a "JOIN" message (GL1 in FIG. 2B) will attempt to form a new group containing the previous members plus all members in the joining group (see FIG. 2C). This is accomplished by sending a "PTC" ("Prepare To Commit")
20 message to all members of the new group.

Nodes receiving a "PTC" message reply with a "PTC_ACK" message as shown in FIG. 2D. All nodes from which a "PTC_ACK" message was received are included in the new group. The group leader (GL1) sends a "COMMIT" message,
25 which contains the entire group membership list, to all new group members. Referring to FIG. 2E, to speed-up the transmission of the "COMMIT" message, a "COMMIT_BCAST" message is sent by the group leader to a small number of

nodes called the "mayors" 12' and each of those will send the "COMMIT" message to all members in a sub-group for which each mayor is responsible (see FIG. 2F). (Because the network is assumed to be unreliable, messages like "PTC",
5 "COMMIT_BCAST", and "COMMIT" are retried a number of times if the acknowledgment is not received.)

Receiving a "COMMIT" message marks the transition to the new group (shown in FIG. 2G), which now contains the old members plus the joining members. After receiving this
10 message, each group member starts sending "HEART BEAT" messages to its (possibly new) downstream neighbor.

When a node is initialized, it forms a singleton adapter group (of which the node is the group leader) in each of its adapters. The node then starts sending and
15 receiving "PROCLAIM" messages.

Referring now to FIGS. 3A-3C, a node will monitor "HEART BEAT" messages 20 (see FIG. 3A) coming from its "upstream neighbor" (the adapter in the group that has the next highest IP address among the group members). When no
20 "HEART BEAT" messages are received for some predefined period of time, the "upstream neighbor" is assumed to have failed. A "DEATH" message is then sent to the group leader requesting that a new group be formed (see FIG. 3B).

Upon receiving a "DEATH" message, the group leader
25 attempts to form a new group containing all adapters in the current group except the adapter that was detected as failed. As shown in FIG. 3C, the group leader sends a "PTC" message to all members of the new group. The protocol then

follows the same sequence as that described above for the JOIN protocol.

5 A node reachability protocol is used to allow computation of the set of nodes that are reachable from a local node (and therefore considered alive). Since not all nodes may be connected to the same network, some nodes may be reachable only through a sequence of multiple network hops. Node reachability can only be computed when information about all networks, even those that do not span
10 all nodes, is taken into account.

15 To compute node reachability, an eventual agreement protocol is used: reachability information at each network is propagated to all networks; when the network topology stops changing, eventually all nodes will have consistent information about all networks. Each node will then be able to compute the set of reachable nodes independently and arrive at a consistent result.

20 Periodically, and until the stopping criteria instruct the daemon to stop doing so, the nodes send the following messages:

25 a "Node Connectivity Message" (NCM or NODE_CONNECTIVITY) is sent from all group members to the GL (see FIG. 4A). A NCM for a given network contains the AMG id for that network plus all the "disabled AMG ids" for the local adapters that are disabled. A node must send NCMs to each GL of the groups to which the local adapters belong.

the GL stores all the information coming from the NCMs in a "Node Connectivity Table" (NCT). The NCT stores the (local view of the) global network topology and contains the AMG id for each node and network adapter in the system. Any two nodes that have the same AMG id are assumed to be connected to each other by the same network.

a "Group Connectivity Message" (GCM or GROUP_CONNECTIVITY) is sent from each GL to all group members (see FIG. 4B). The GCM contains the AMG id and the list of nodes that belong to the AMG. Also, for each of these nodes, a list of all "disabled AMG ids" (in the other networks) is included. The information needed to send the GCM is extracted from the GL's NCT.

a node that receives a GCM updates its own NCT with the information in the message. If a daemon receiving a GCM notices that there are some groups to which the local adapters belong, whose members will not have received that GCM, the daemon forwards the GCM to these groups (reference node 2 in FIG. 4C). The goal is to propagate the GCM to all the nodes in the system, even those that are not directly connected to the network that originated the GCM.

In FIG. 4C, the GCM for AMG_1 is forwarded by either node 2 or node 3 to nodes 4 and 5 through network2.

Notice that the information sent in an NCM and GCM is a subset of the sender's NCT.

In accordance with the present invention, a node can stop sending NCMs for a given network if the corresponding GCM sent by the GL already reflects the information sent from that node to the GL in a previous NCM. This is done by
5 comparing the information sent in the last NCM with the information in the incoming GCM that refers to the local node.

The sending of NCMs in all groups is resumed when the GCM information conflicts with the local information, or
10 when the daemon detects that a new AMG id is in place for some network to which a local adapter is connected (the latter can be detected by comparing the information in the GCM with that stored in the NCT). NCMs are also resumed when a local adapter is detected as disabled.

A node may stop sending GCMs after a fixed number of them have been sent, because it is assumed that at least some of them will have arrived at all the (live) nodes in the system. Sending of GCMs is resumed by a GL when a new AMG id is formed, which happens when a new adapter joins the
20 group or an existing member is expelled from it. To allow recently powered up nodes to obtain all the needed GCMs, a node will also resume sending GCMs (for a fixed number of times) when it receives any GCM or NCM that conflicts with the receiving node's NCT. In addition, GCMs are resumed by
25 a node when one of its adapters is moved to the "disabled" state.

This mechanism (illustrated in one example in FIGS. 6A-6H) works in the following way:

- a node is powered up, and its daemon is started;
 - the node's adapters join a number of AMGs;
 - GCMs are sent for the newly formed AMGs by their respective GLs;
- 5 - GCMs reach all the live nodes either directly or by using the GCM forwarding mechanism;
- all GLs receiving the new GCM resume sending their GCMs, since the arriving GCM includes information about a newly formed group and thus causes a
- 10 change in the NCT's contents;
- the recently powered up node obtains GCMs from all the groups.

It is assumed that if all nodes are up then all will get at least one of the GCMs sent by a GL (and forwarded to

15 other networks as needed). If a node is not up at this point, it will get the GCM later on when it is powered on, since the resuming criteria are applied when the node becomes alive.

Both NCMs and GCMs are resumed at a node when any of

20 its adapters becomes disabled. This is consistent with the strategy of resuming GCMs when a node perceives changes in topology.

The following reasoning explains why at least one of the GCMs should reach all nodes with high probability. If

25 no GCMs reach a node, even after several tries, this usually points to an existing network problem. However, since adapters in an AMG are supposed to monitor each other, network problems should be detected well before all GCMs are sent. The detection of network problems should result in

new AMGs formed by the adapters that can communicate with each other. As a result, the new GCMs will flow through adapters that are known to be working.

FIGS. 5A-5E depict one example of topology propagation in accordance with the present invention. In FIG. 5A, the distributed computing environment is shown to include nodes 1-6 and networks 1 & 2 which have AMG A_1 and AMG B_1, respectively. Each node of the environment has the correct global topology configuration in a respective NCT. For example, reference FIG. 5B wherein the NCT at node 5 is shown.

In FIG. 5C, node 2 is assumed to disappear resulting in a new adapter membership group (AMG A_2) being created by nodes 1, 3 & 4. At the time of creation of AMG A_2, the NCT at node 5, which is shown in FIG. 5D, has yet to reflect the disappearance of node 2 from the computing environment.

Node 5 becomes aware of the disappearance of node 2 by group leader GL1 forwarding a group connectivity message (GCM) to nodes 1 & 3 of AMG A_2. Node 3, which has local adapters to both network 1 and network 2, then forwards the transmitted GCM to nodes 5 & 6 of AMG B_1. As noted above, the forwarding of the GCM could be accomplished by either node 3 or node 4 since both nodes are common to both networks. FIG. 5F depicts the updated NCT at node 5 upon receipt of the forwarded GCM. Note that node 2 becomes isolated from node 5 in that it remains a member of A_1 which is unreachable by node 5 through any hopping from A_2 or B_1.

FIGS. 6A-6H depict another example of topology propagation in accordance with the principles of the present invention. In this example, node 2 is to become active within the distributed computing environment depicted in FIG. 6A. In this environment, nodes 1, 3 & 4 belong to AMG A_1, while nodes 3, 4, 5 & 6 belong to AMG B_1. FIG. 6B depicts the NCT at node 5, and the NCT at node 2 for the distributed computing environment of FIG. 6A.

In FIG. 6C, node 2 is now alive and a new adapter membership group, (AMG A_2), has been formed. At this point in time, the topology configuration in NCT at node 5 and NCT at node 2 is shown in FIG. 6D, which is the same as that of FIG. 6B.

The nodes are informed of the new AMG by GL1 forwarding a group connectivity message (GCM) to nodes 1, 2 & 3, and by node 3 forwarding the GCM to nodes 5 & 6 as shown in FIG. 6E. Upon receipt of the GCM, each node updates its NCT, resulting in the NCT at node 5 and NCT at node 2 shown in FIG. 6F. To complete the topology update, group leader 2 of AMG B_1 responds to the new information by sending its own GCM, which advises node 2 of AMG B_1. As shown in FIG. 6G, the GCM from GL2 is sent to nodes 3, 4 & 5, with node 3 forwarding the message along to node 1 & node 2 of AMG A_2. The updated topology information in NCT at node 5 and NCT at node 2 is shown in FIG. 6H.

FIG. 7 depicts a state diagram for a group leader implementing topology propagation in accordance with this invention. In state 1, the group leader is sending GCM messages to the nodes in its group. Upon occurrence of a

predefined event, for example, of a message count reaching a preset limit, the group leader enters a second state where it is not sending GCM messages to the members of its group. Thereafter, the group leader remains in state 2 until there
5 is a change in the distributed computing environment. Specifically, the group leader transitions to state 1 if:
(1) the group leader receives an NCM which conflicts with a local NCT; (2) the group leader receives a GCM which conflicts with its local NCT; (3) a local adapter of the
10 group leader belongs to a different AMG; or (4) a local adapter of the group leader is considered disabled.

Those skilled in the art will note from the above description that presented herein is a mechanism to stop and restart sending of topology propagation messages within a
15 distributed computing environment. This mechanism obviates the need to send network topology information periodically to the nodes in the distributed system. Once the topology stops changing, all GCMs in the system will stop within a finite amount of time. A mechanism in accordance with the
20 present invention is used by topology services to disseminate topology information among all nodes in the system. The NCT is used by topology services to:

- Compute the set of nodes that are reachable from the local node.
- 25 • Compute the route to each reachable node. The route is used by reliable messaging (PRM) to "source-route" packets to destinations.

The present invention can be included, for example, in an article of manufacture (e.g., one or more computer

program products) having, for instance, computer usable media. This media has embodied therein, for instance, computer readable program code means for providing and facilitating the capabilities of the present invention. The
5 articles of manufacture can be included as part of the computer system or sold separately.

Additionally, at least one program storage device readable by machine, tangibly embodying at least one program of instructions executable by the machine, to perform the
10 capabilities of the present invention, can be provided.

The flow diagrams depicted herein are provided by way of example. There may be variations to these diagrams or the steps (or operations) described herein without departing from the spirit of the invention. For instance, in certain
15 cases, the steps may be performed in differing order, or steps may be added, deleted or modified. All of these variations are considered to comprise part of the present invention as recited in the appended claims.

While the invention has been described in detail herein
20 in accordance with certain preferred embodiments thereof, many modifications and changes therein may be effected by those skilled in the art. Accordingly, it is intended by the appended claims to cover all such modifications and changes as fall within the true spirit and scope of the
25 invention.

Claims

1 1. A method of topology propagation in a distributed
2 computing environment, said method comprising:

3 sending group connectivity messages from at least
4 one group leader to identified nodes of at least one
5 group of nodes within the distributed computing
6 environment;

7 discontinuing said sending of group connectivity
8 messages during a time period of no topology change
9 within the distributed computing environment; and

10 reinitiating sending of group connectivity
11 messages from the at least one group leader upon
12 identification of a topology change within the
13 distributed computing environment.

1 2. The method of claim 1, wherein the distributed
2 computing environment comprises at least two networks each
3 having at least one group of identified nodes, and wherein
4 said method further comprises employing within each group of
5 the at least two networks a heartbeat protocol to ensure
6 continued presence of each identified node within the group.

1 3. The method of claim 2, wherein the at least two
2 networks of the distributed computing environment comprise
3 heterogenous networks.

1 4. The method of claim 2, wherein at least one node
2 of the distributed computing environment has at least two
3 adapters, said at least two adapters coupling said at least
4 one node to said at least two networks, and wherein said
5 sending comprises sending first group connectivity messages
6 (GCMs) from a first group leader to identified nodes of a
7 first group of nodes on a first network of said at least two
8 networks, said at least one node comprising an identified
9 node of said first group of nodes, and forwarding said first
10 GCMs by said at least one node to a second group of nodes on
11 a second network of said at least two networks.

1 5. The method of claim 4, wherein said first GCMs
2 received at identified nodes of said first group of nodes
3 and identified nodes of said second group of nodes are
4 employed by each said identified node to update a local
5 network connectivity table (NCT).

1 6. The method of claim 4, wherein said sending
2 further comprises sending second GCMs from a second group
3 leader to identified nodes of the second group of nodes, and
4 forwarding said second GCMs by said at least one node to the
5 first group of nodes on the first network of the at least
6 two networks.

1 7. The method of claim 6, wherein said sending second
2 GCMs by said second group leader is responsive to receiving
3 new information in said forwarded first GCMs at said second
4 group leader.

1 8. The method of claim 6, wherein said discontinuing
2 comprises for each group leader discontinuing said sending
3 of group connectivity messages when a number of messages
4 sent from the group leader reaches a set limit after
5 identification by said group leader of a topology change
6 within the distributed computing environment.

1 9. The method of claim 8, wherein said reinitiating
2 comprises identifying said topology change within a
3 distributed computing environment, said identifying
4 comprising at least one of: receiving at a group leader a
5 node connectivity message which conflicts with a local
6 network connectivity table value, receiving at a group
7 leader a group connectivity message which conflicts with a
8 local network connectivity table value, identifying that a
9 local adapter belongs to a different adapter membership
10 group, or identifying that a local adapter has become
11 disabled.

1 10. The method of claim 1, wherein said discontinuing
2 comprises for each group leader discontinuing said sending
3 of group connectivity messages when a number of messages
4 sent from the group leader reaches a set limit after
5 identification of the topology change within the distributed
6 computing environment.

1 11. The method of claim 1, further comprising
2 implementing said sending, said discontinuing, and said
3 reinitiating without employing acknowledgment messages
4 during said topology propagation.

1 12. The method of claim 1, wherein said reinitiating
2 sending of group connectivity messages comprises at least
3 one of receiving at a group leader a node connectivity
4 message which conflicts with a local network connectivity
5 table value, receiving at a group leader a group
6 connectivity message which conflicts with a local network
7 connectivity table value, identifying that a local adapter
8 belongs to a different adapter membership group, or
9 identifying that a local adapter has become disabled.

1 13. A system for topology propagation in a distributed
2 computing environment, said system comprising:

3 means for sending group connectivity messages from
4 at least one group leader to identified nodes of at
5 least one group of nodes within the distributed
6 computing environment;

7 means for discontinuing said sending of group
8 connectivity messages during a time period of no
9 topology change within the distributed computing
10 environment; and

11 means for reinitiating sending of group
12 connectivity messages from the at least one group
13 leader upon identification of a topology change within
14 the distributed computing environment.

1 14. The system of claim 13, wherein the distributed
2 computing environment comprises at least two networks each
3 having at least one group of identified nodes, and wherein
4 said system further comprises means for employing within
5 each group of the at least two networks a heartbeat protocol
6 to ensure continued presence of each identified node within
7 the group.

1 15. The system of claim 14, wherein the at least two
2 networks of the distributed computing environment comprise
3 heterogenous networks.

1 16. The system of claim 14, wherein at least one node
2 of the distributed computing environment has at least two
3 adapters, said at least two adapters coupling said at least
4 one node to said at least two networks, and wherein said
5 means for sending comprises means for sending first group
6 connectivity messages (GCMs) from a first group leader to
7 identified nodes of a first group of nodes on a first
8 network of said at least two networks, said at least one
9 node comprising an identified node of said first group of
10 nodes, and means for forwarding said first GCMs by said at
11 least one node to a second group of nodes on a second
12 network of said at least two networks.

1 17. The system of claim 16, wherein said first GCMs
2 received at identified nodes of said first group of nodes
3 and identified nodes of said second group of nodes are
4 employed by each said identified node to update a local
5 network connectivity table (NCT).

1 18. The system of claim 16, wherein said means for
2 sending further comprises means for sending second GCMs from
3 a second group leader to identified nodes of the second
4 group of nodes, and means for forwarding said second GCMs by
5 said at least one node to the first group of nodes on the
6 first network of the at least two networks.

1 19. The system of claim 18, wherein said means for
2 sending second GCMs by said second group leader is
3 responsive to receiving new information in said forwarded
4 first GCMs at said second group leader.

1 20. The system of claim 18, wherein said means for
2 discontinuing comprises for each group leader means for
3 discontinuing said sending of group connectivity messages
4 when a number of messages sent from the group leader reaches
5 a set limit after identification by said group leader of a
6 topology change within the distributed computing
7 environment.

1 21. The system of claim 20, wherein said means for
2 reinitiating comprises means for identifying said topology
3 change within a distributed computing environment, said
4 means for identifying being responsive to at least one of:
5 receiving at a group leader a node connectivity message
6 which conflicts with a local network connectivity table
7 value, receiving at a group leader a group connectivity
8 message which conflicts with a local network connectivity
9 table value, identifying that a local adapter belongs to a
10 different adapter membership group, or identifying that a
11 local adapter has become disabled.

1 22. The system of claim 13, wherein said means for
2 discontinuing comprises for each group leader means for
3 discontinuing said sending of group connectivity messages
4 when a number of messages sent from the group leader reaches
5 a set limit after identification of the topology change
6 within the distributed computing environment.

1 23. The system of claim 13, wherein said means for
2 sending, said means for discontinuing, and said means for
3 reinitiating are implemented without employing
4 acknowledgment messages during said topology propagation.

1 24. The system of claim 13, wherein said means for
2 reinitiating sending of group connectivity messages is
3 responsive to at least one of receiving at a group leader a
4 node connectivity message which conflicts with a local
5 network connectivity table value, receiving at a group
6 leader a group connectivity message which conflicts with a
7 local network connectivity table value, identifying that a
8 local adapter belongs to a different adapter membership
9 group, or identifying that a local adapter has become
10 disabled.

1 25. At least one program storage device readable by a
2 machine tangibly embodying at least one program of
3 instructions executable by the machine to perform a method
4 of topology propagation in a distributed computing
5 environment, comprising:

6 sending group connectivity messages from at least
7 one group leader to identified nodes of at least one
8 group of nodes within the distributed computing
9 environment;

10 discontinuing said sending of group connectivity
11 messages during a time period of no topology change
12 within the distributed computing environment; and

13 reinitiating sending of group connectivity
14 messages from the at least one group leader upon
15 identification of a topology change within the
16 distributed computing environment.

1 26. The at least one program storage device of claim
2 25, wherein the distributed computing environment comprises
3 at least two networks each having at least one group of
4 identified nodes, and wherein said method further comprises
5 employing within each group of the at least two networks a
6 heartbeat protocol to ensure continued presence of each
7 identified node within the group.

1 27. The at least one program storage device of claim
2 26, wherein the at least two networks of the distributed
3 computing environment comprise heterogenous networks.

1 28. The at least one program storage device of claim
2 26, wherein at least one node of the distributed computing
3 environment has at least two adapters, said at least two
4 adapters coupling said at least one node to said at least
5 two networks, and wherein said sending comprises sending
6 first group connectivity messages (GCMs) from a first group
7 leader to identified nodes of a first group of nodes on a
8 first network of said at least two networks, said at least
9 one node comprising an identified node of said first group
10 of nodes, and forwarding said first GCMs by said at least
11 one node to a second group of nodes on a second network of
12 said at least two networks.

1 29. The at least one program storage device of claim
2 28, wherein said first GCMs received at identified nodes of
3 said first group of nodes and identified nodes of said
4 second group of nodes are employed by each said identified
5 node to update a local network connectivity table (NCT).

1 30. The at least one program storage device of claim
2 28, wherein said sending further comprises sending second
3 GCMs from a second group leader to identified nodes of the
4 second group of nodes, and forwarding said second GCMs by
5 said at least one node to the first group of nodes on the
6 first network of the at least two networks.

1 31. The at least one program storage device of claim
2 30, wherein said sending second GCMs by said second group
3 leader is responsive to receiving new information in said
4 forwarded first GCMs at said second group leader.

1 32. The at least one program storage device of claim
2 30, wherein said discontinuing comprises for each group
3 leader discontinuing said sending of group connectivity
4 messages when a number of messages sent from the group
5 leader reaches a set limit after identification by said
6 group leader of a topology change within the distributed
7 computing environment.

1 33. The at least one program storage device of claim
2 32, wherein said reinitiating comprises identifying said
3 topology change within a distributed computing environment,
4 said identifying comprising at least one of: receiving at a
5 group leader a node connectivity message which conflicts
6 with a local network connectivity table value, receiving at
7 a group leader a group connectivity message which conflicts
8 with a local network connectivity table value, identifying
9 that a local adapter belongs to a different adapter
10 membership group, or identifying that a local adapter has
11 become disabled.

1 34. The at least one program storage device of claim
2 25, wherein said discontinuing comprises for each group
3 leader discontinuing said sending of group connectivity
4 messages when a number of messages sent from the group
5 leader reaches a set limit after identification of the
6 topology change within the distributed computing
7 environment.

1 35. The at least one program storage device of claim
2 25, further comprising implementing said sending, said
3 discontinuing, and said reinitiating without employing
4 acknowledgment messages during said topology propagation.

1 36. The at least one program storage device of claim
2 25, wherein said reinitiating sending of group connectivity
3 messages comprises at least one of receiving at a group
4 leader a node connectivity message which conflicts with a
5 local network connectivity table value, receiving at a group
6 leader a group connectivity message which conflicts with a
7 local network connectivity table value, identifying that a
8 local adapter belongs to a different adapter membership
9 group, or identifying that a local adapter has become
10 disabled.

* * * * *

TOPOLOGY PROPAGATION IN A DISTRIBUTED
COMPUTING ENVIRONMENT WITH NO TOPOLOGY
MESSAGE TRAFFIC IN STEADY STATE

Abstract of the Disclosure

5 A topology propagation facility is provided for
maintaining a common network topology database at different
nodes in a distributed computing environment. The facility
generates no message traffic when the distributed computing
environment is in steady state. This is accomplished by
10 discontinuing sending of group connectivity messages during
a time period of no topology change within the distributed
environment. Sending of group connectivity messages is
reinitiated from at least one group leader upon
identification by the group leader of at least one topology
15 change within the distributed computing environment. Group
connectivity messages are forwarded from one group of nodes
on a first network to another group of nodes on a second
network using a node common to both groups of nodes. The
networks of the distributed computing environment can
20 comprise heterogenous networks such that the topology
propagation facility presented facilitates interoperability
of the networks.

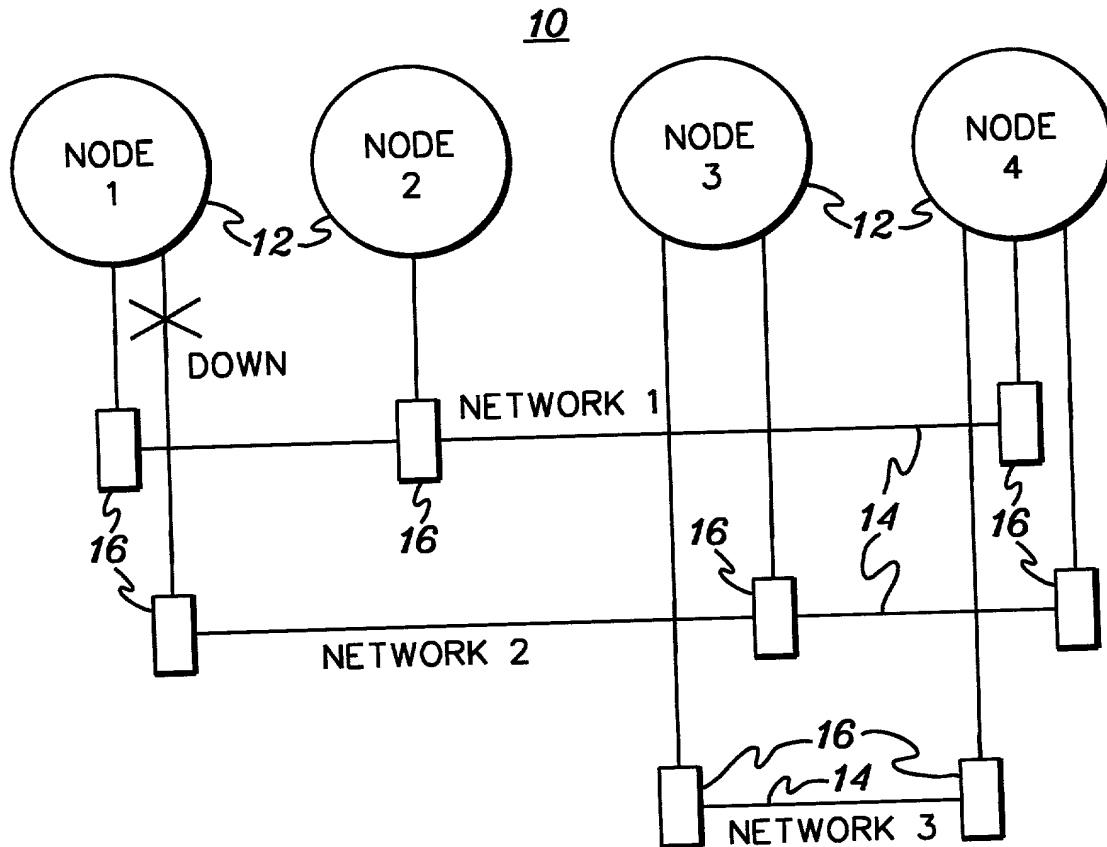


fig. 1

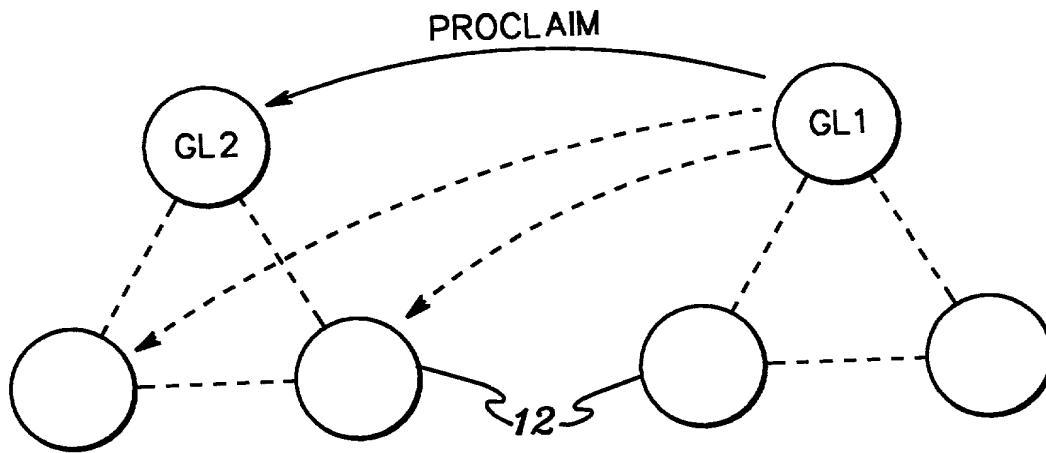


fig. 2A

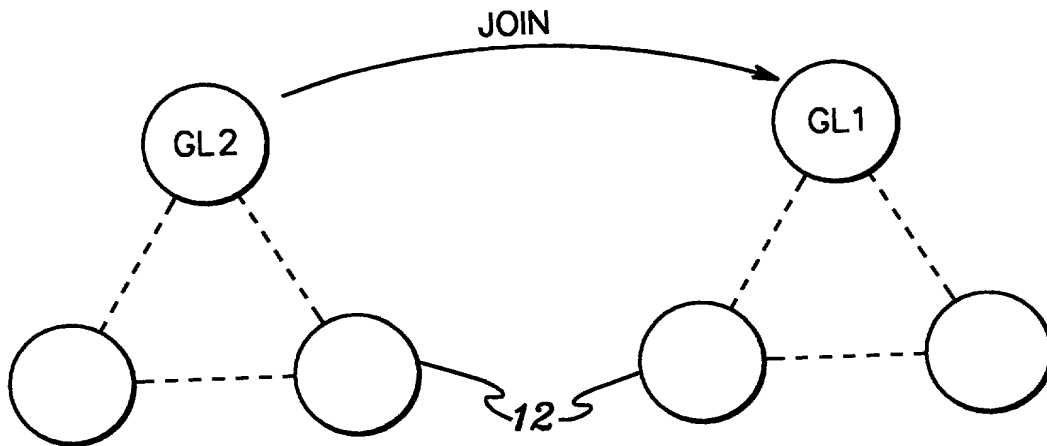


fig. 2B

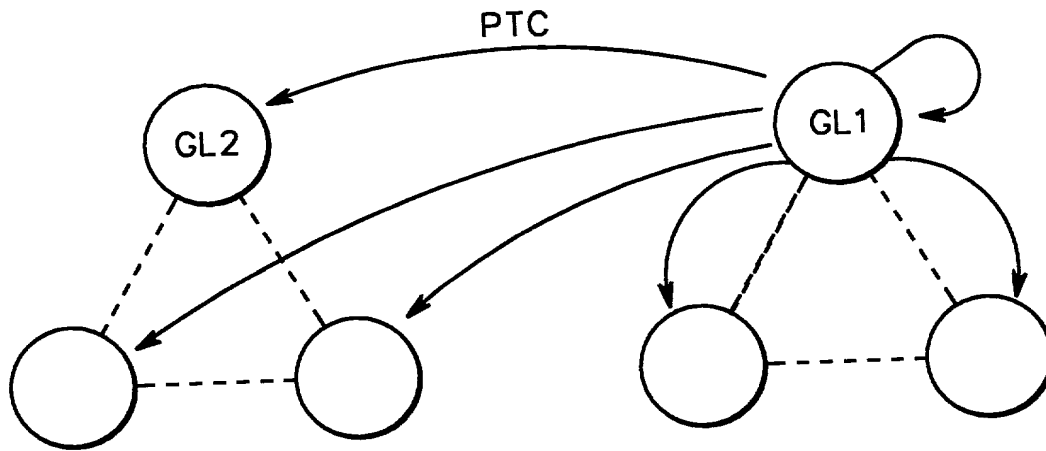


fig. 2C

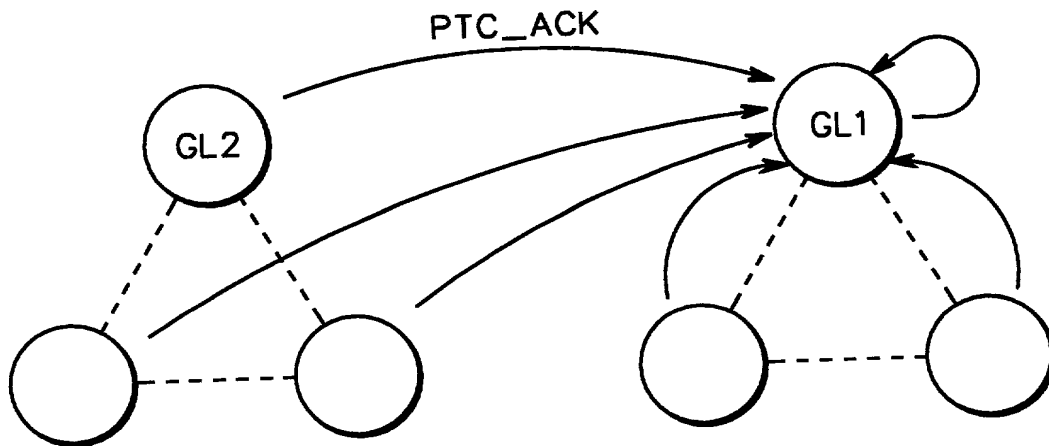


fig. 2D

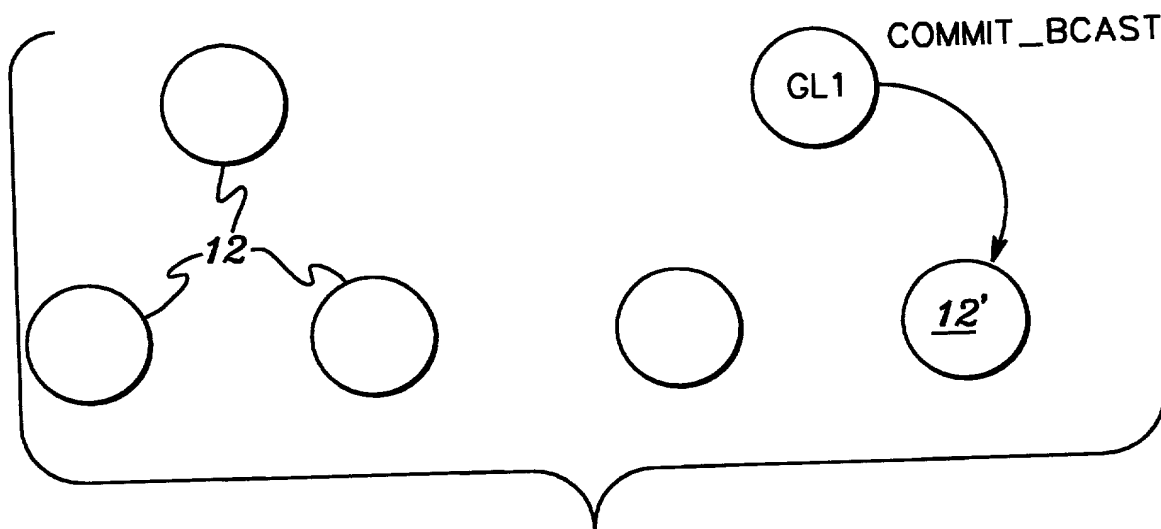


fig. 2E

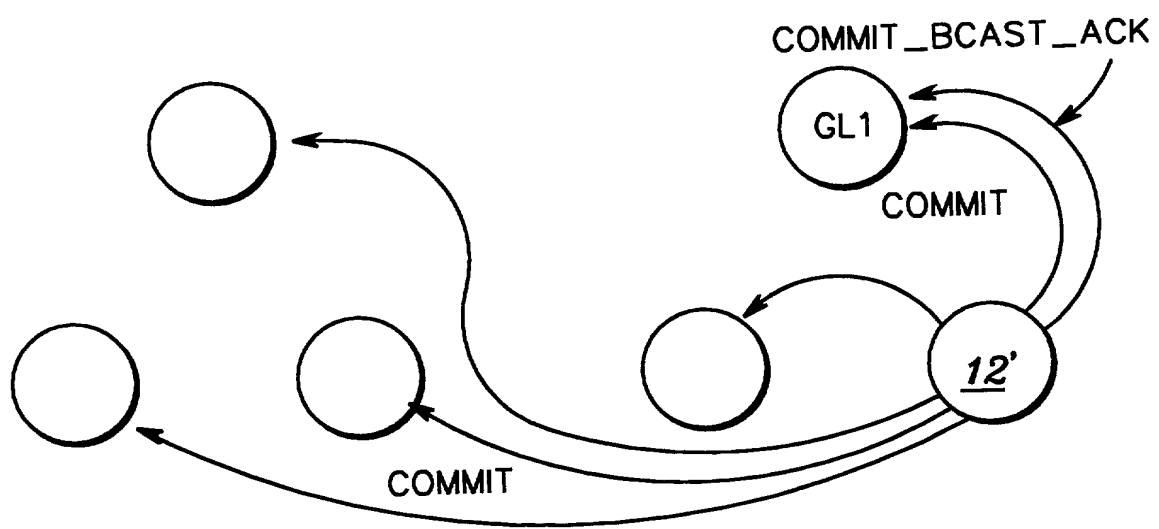


fig. 2F

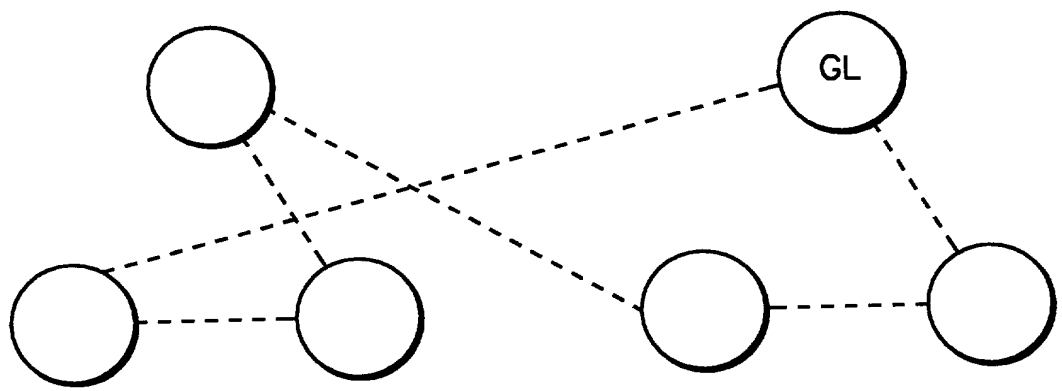


fig. 2G

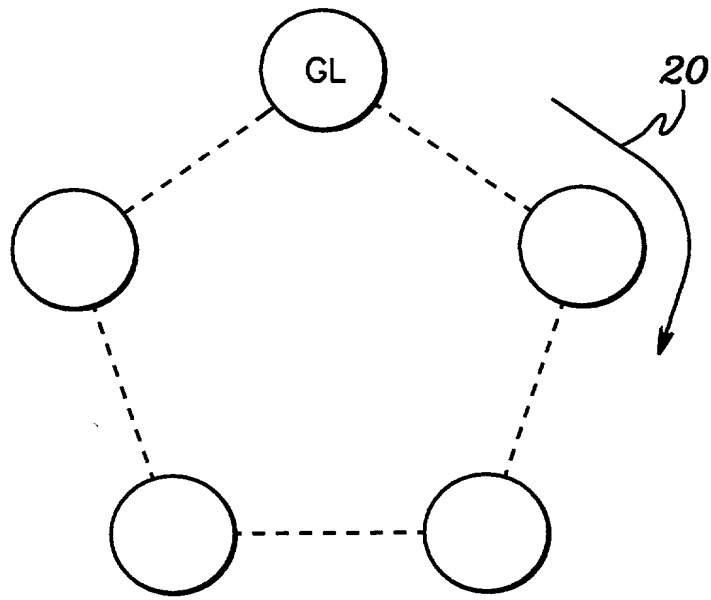


fig. 3A

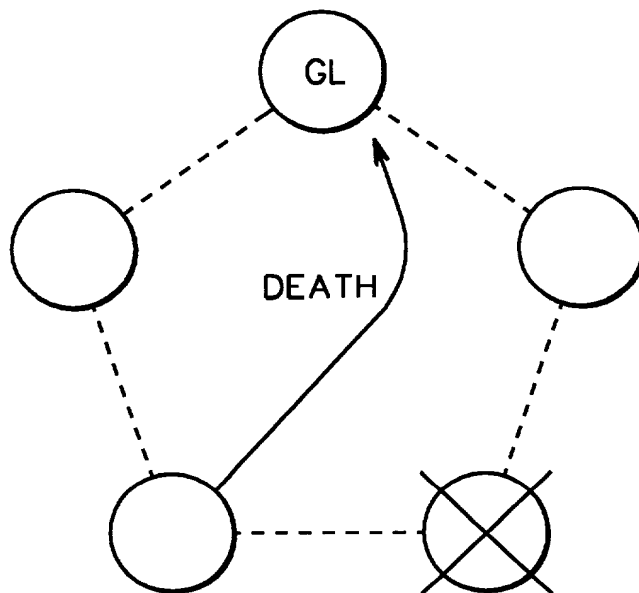


fig. 3B

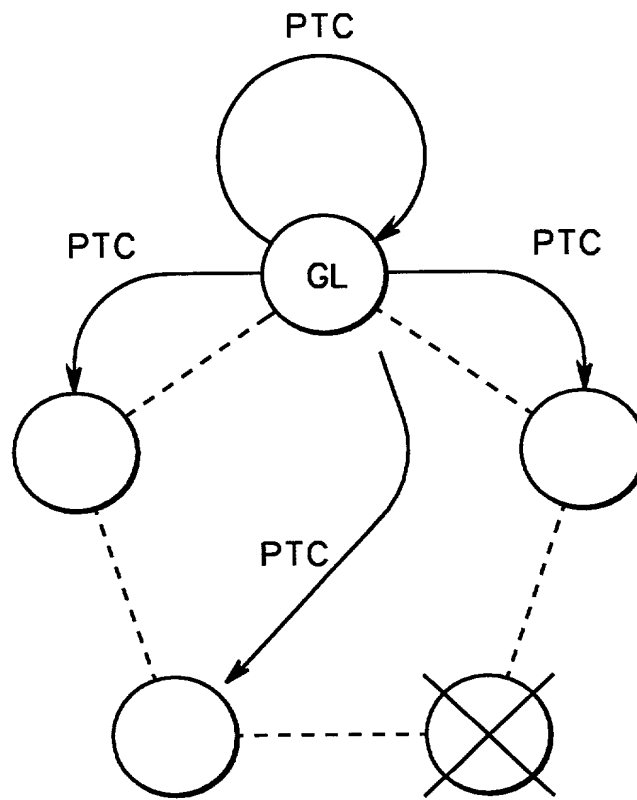


fig. 3C

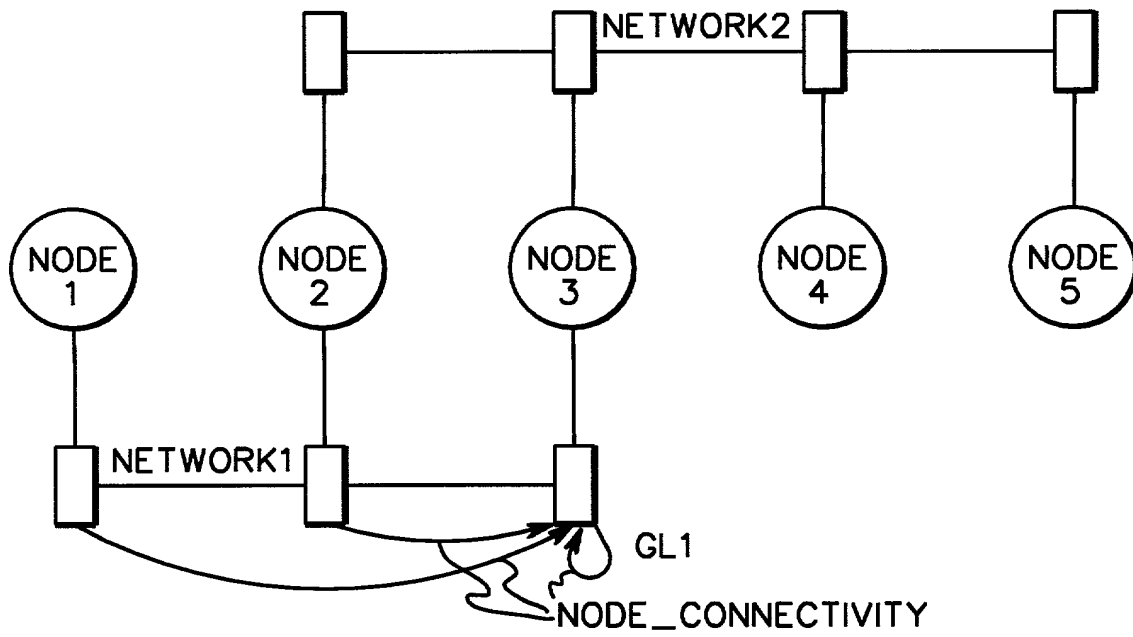


fig. 4A

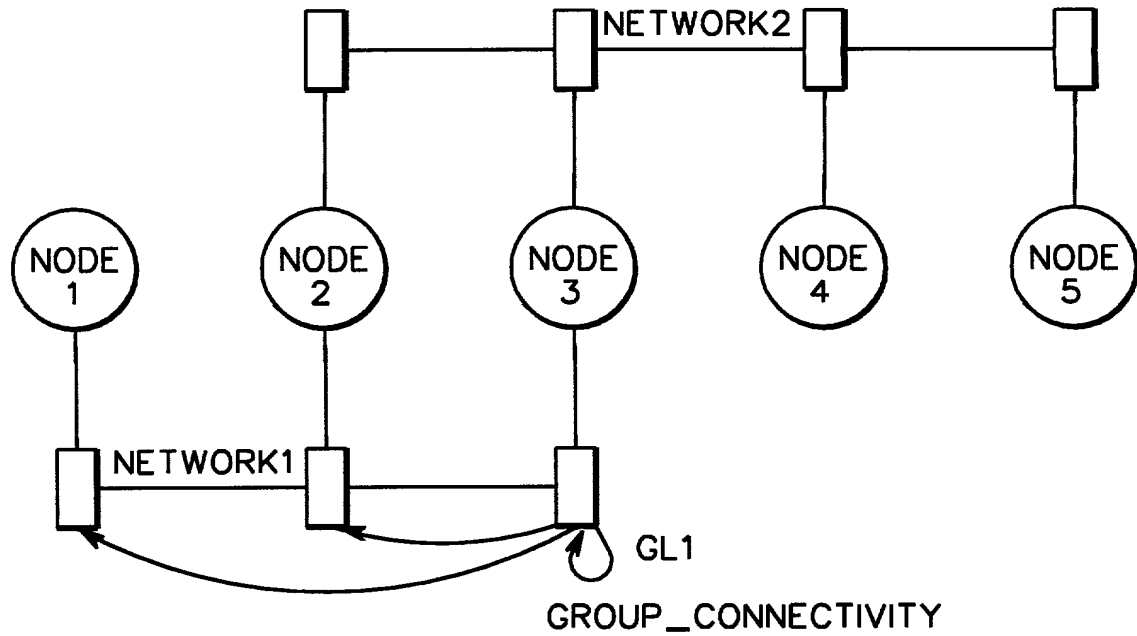


fig. 4B

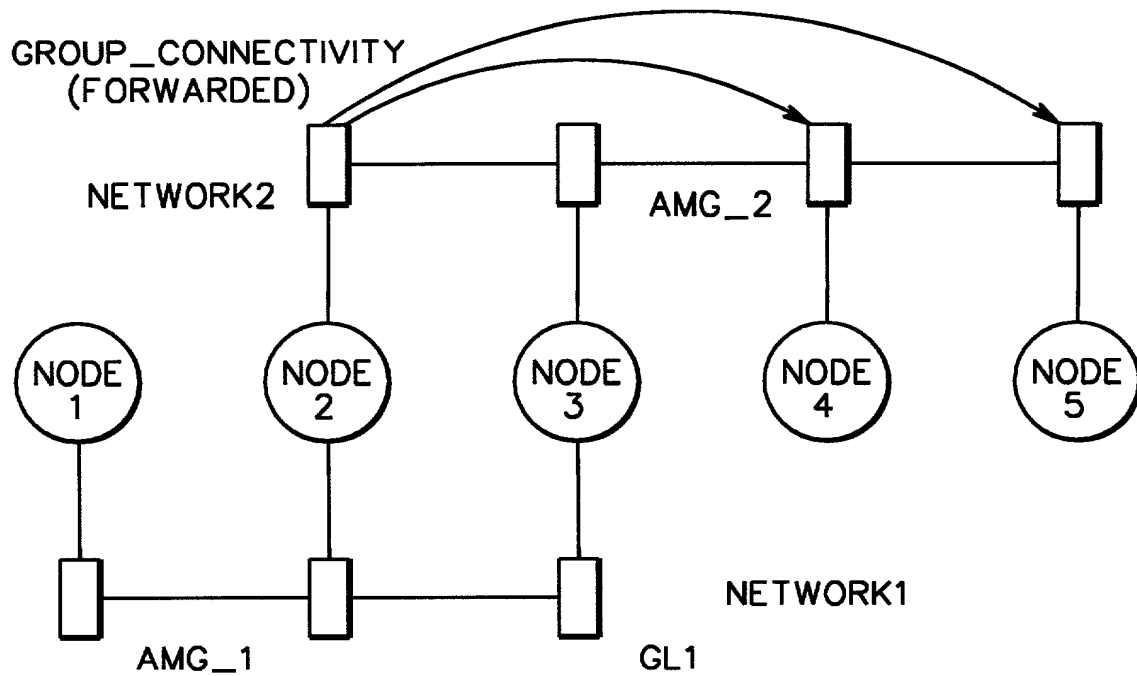


fig. 4C

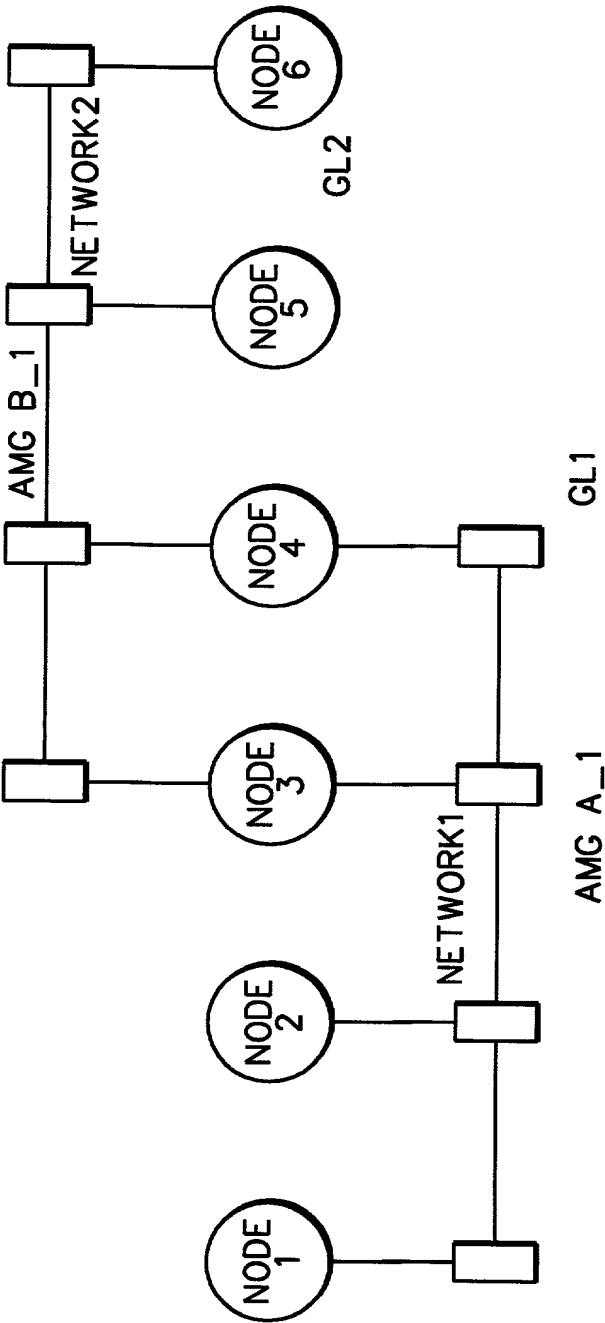


fig. 5A

NCT AT NODE 5

1	2	3	4	5	6
		B_1	B_1	B_1	B_1
A_1	A_1	A_1	A_1		

fig. 5B

FIG. 5C is a schematic diagram of a network topology. The network consists of six nodes, labeled NODE 1 through NODE 6. NODE 1 is a solid circle at the top left. A solid line connects NODE 1 to a solid square below it. This square is connected to a dashed square below it. A dashed line connects the dashed square to a dashed circle labeled ~~NODE 2~~. A solid line connects the dashed square to a solid square to its right. This solid square is connected to a solid circle labeled NODE 3. A solid line connects NODE 3 to another solid square to its right. This square is connected to a solid circle labeled NODE 4. A solid line connects NODE 4 to another solid square to its right. This square is connected to a solid circle labeled NODE 5. A solid line connects NODE 5 to a solid square to its right. This square is connected to a solid circle labeled NODE 6. The label 'AMG A_2' is placed below the solid square connected to NODE 3. The label 'GL1' is placed to the right of the solid square connected to NODE 4. The label 'AMG B_1' is placed above the solid square connected to NODE 5. The label 'GL2' is placed to the right of the solid square connected to NODE 6.

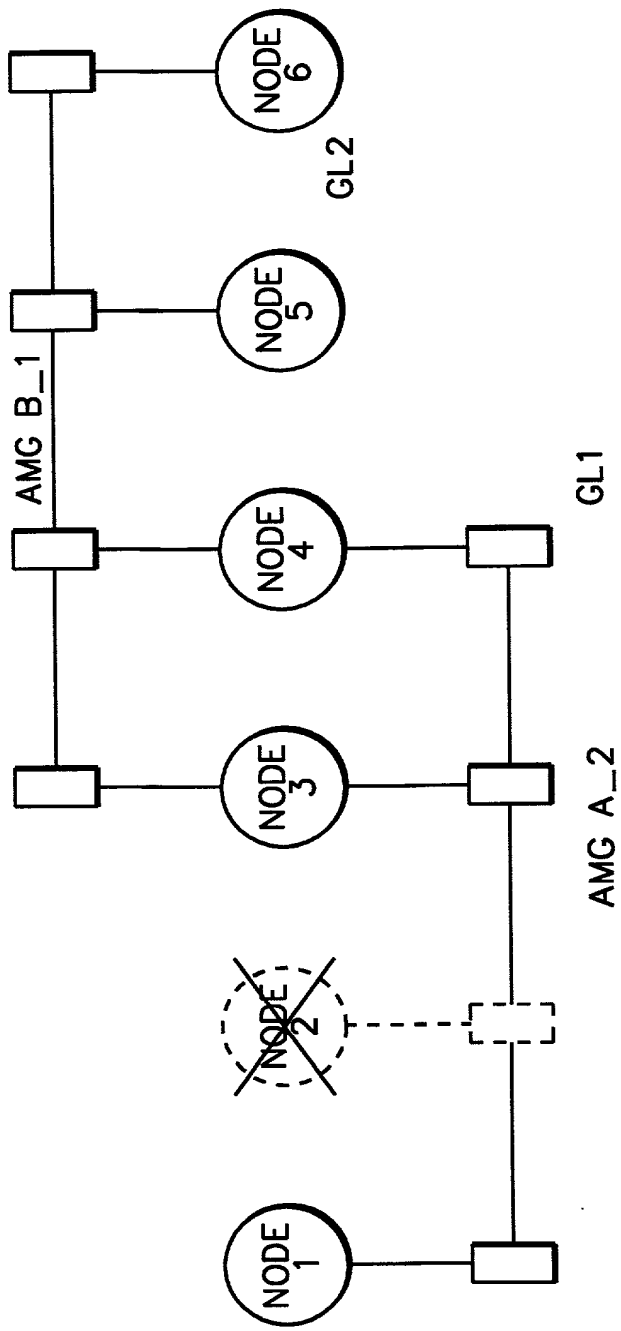


fig. 5C

NCT AT NODE 5

1	2	3	4	5	6
		B_1	B_1	B_1	B_1
A_1	A_1	A_1	A_1		

fig. 5D

12/17
POU9-2000-0017-US1

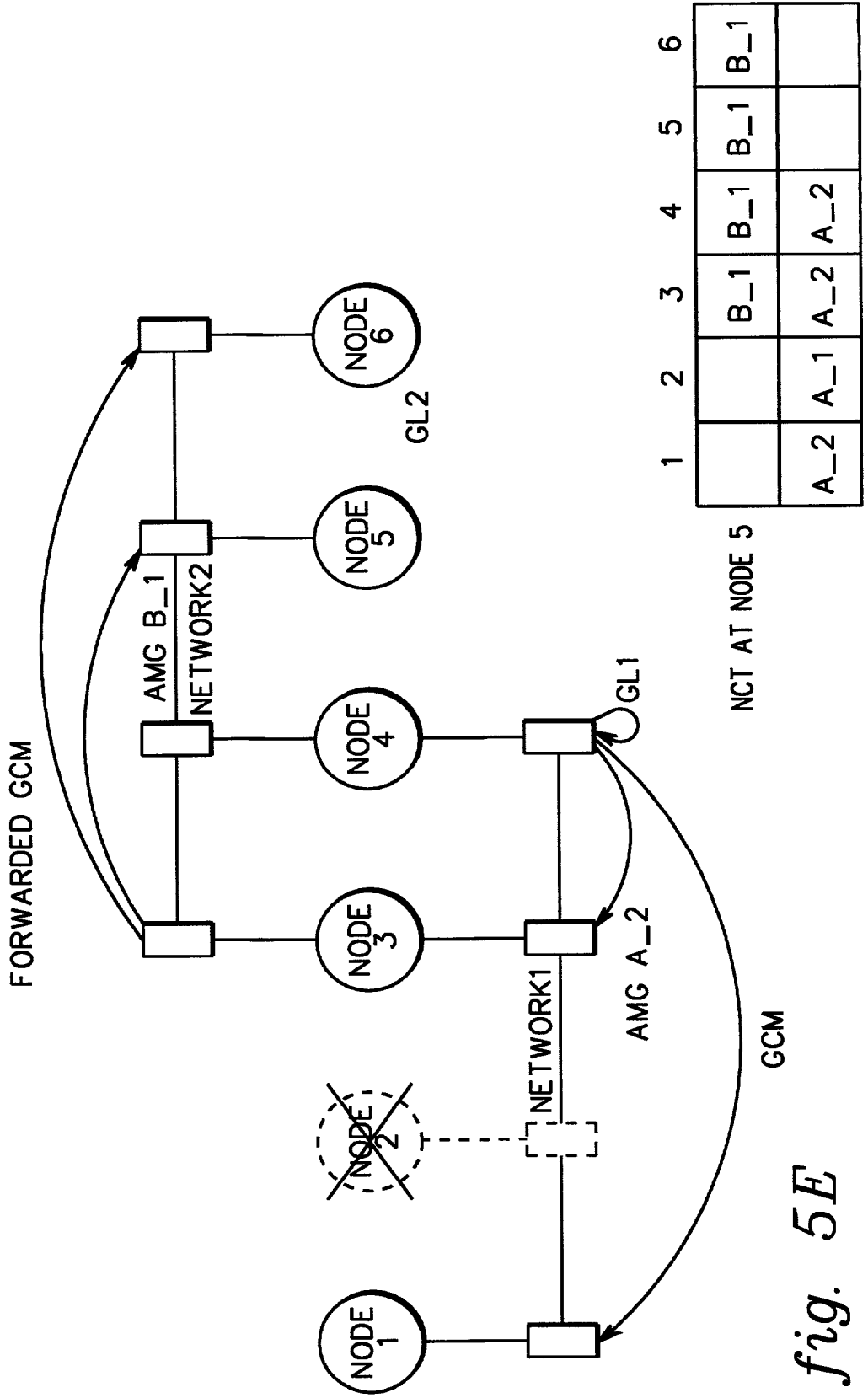


fig. 5E

fig. 5F

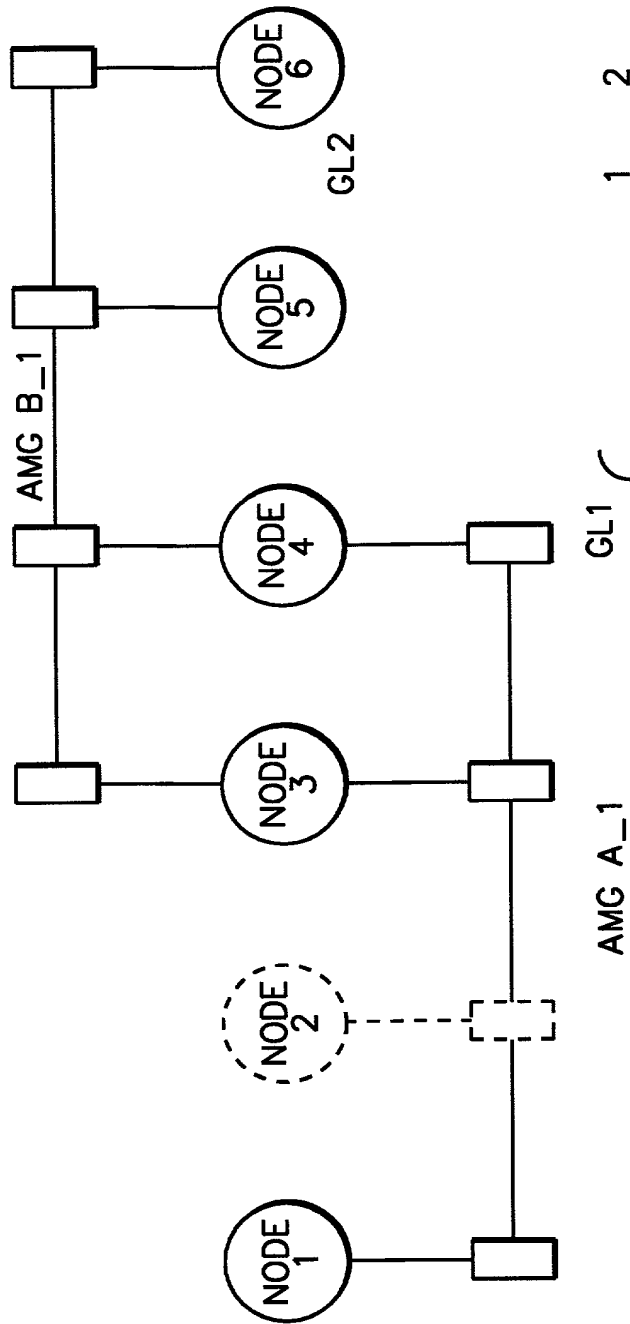


fig. 6A

fig. 6B

	1	2	3	4	5	6
NCT AT NODE 5			B_1	B_1	B_1	B_1
	A_1		A_1	A_1		

	1	2	3	4	5	6
NCT AT NODE 2						

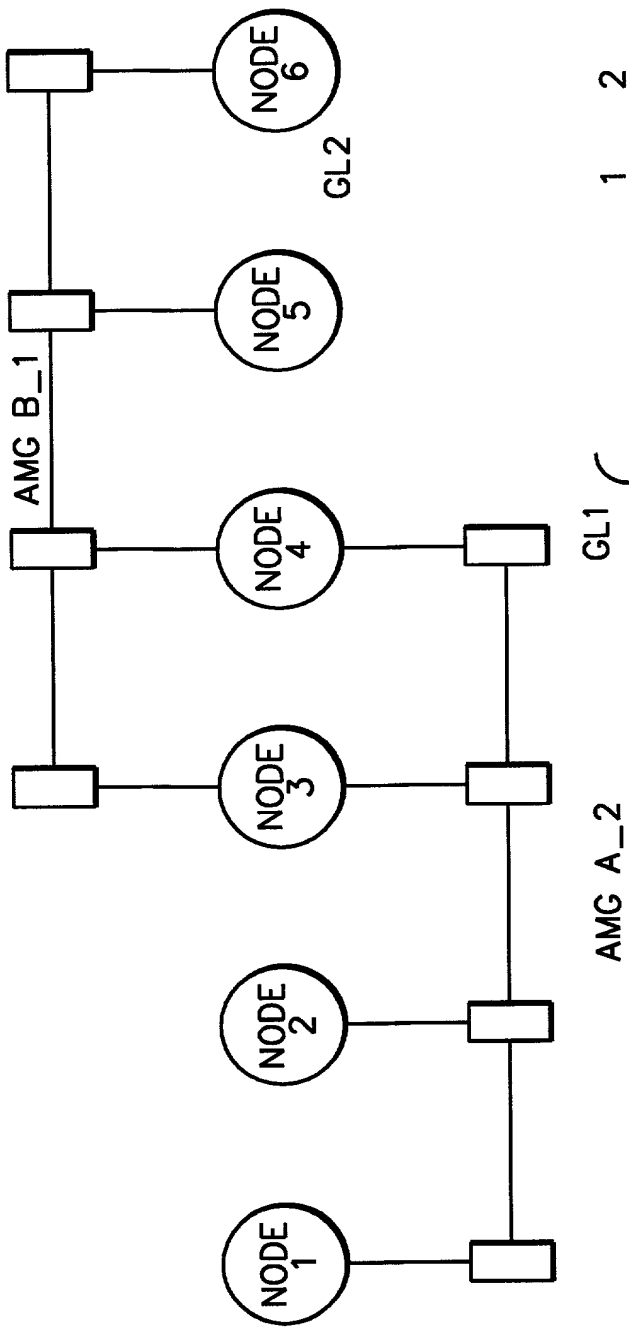


fig. 6C

NCT AT NODE 5					
1	2	3	4	5	6
		B_1	B_1	B_1	B_1
A_1		A_1	A_1		

NCT AT NODE 2					
1	2	3	4	5	6

fig. 6D

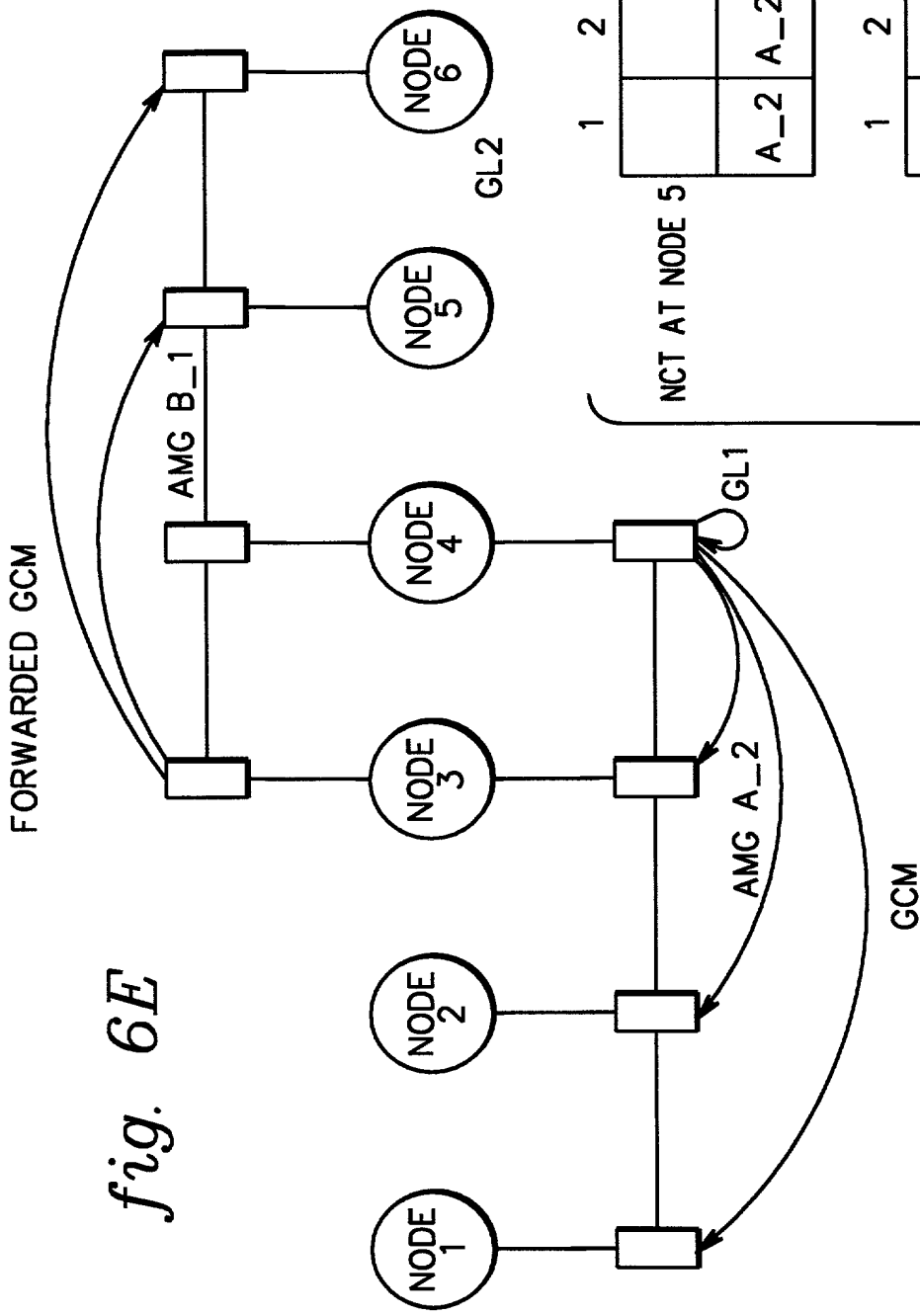


fig. 6E

NCT AT NODE 5

1	2	3	4	5	6
		B_1	B_1	B_1	B_1
A_2	A_2	A_2	A_2		

NCT AT NODE 2

1	2	3	4	5	6
A_2	A_2	A_2	A_2		

fig. 6F

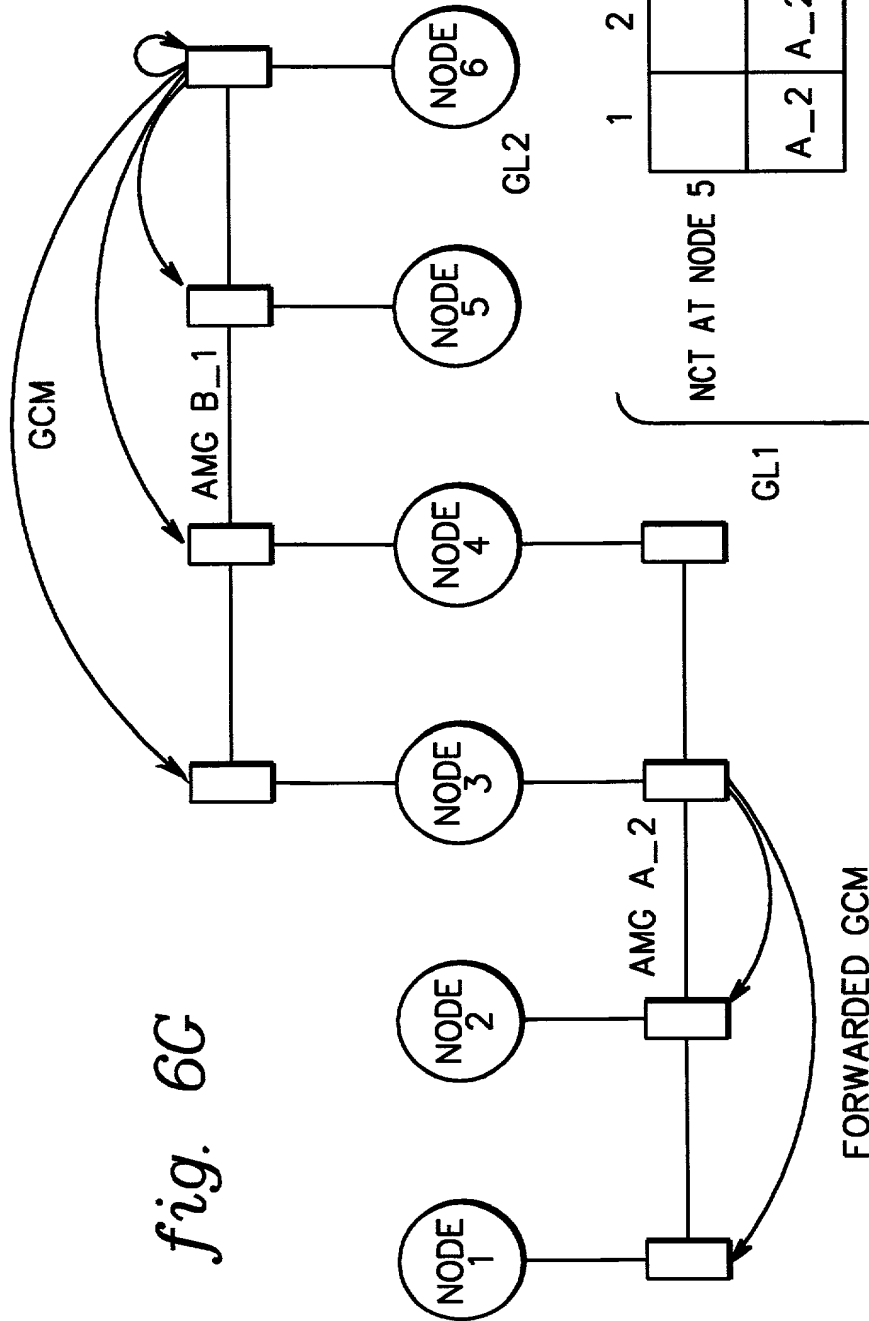


fig. 6H

NCT AT NODE 5

	1	2	3	4	5	6
			B_1	B_1	B_1	B_1
A_2	A_2	A_2	A_2	A_2		

NCT AT NODE 2

	1	2	3	4	5	6
			B_1	B_1	B_1	B_1
A_2	A_2	A_2	A_2	A_2		

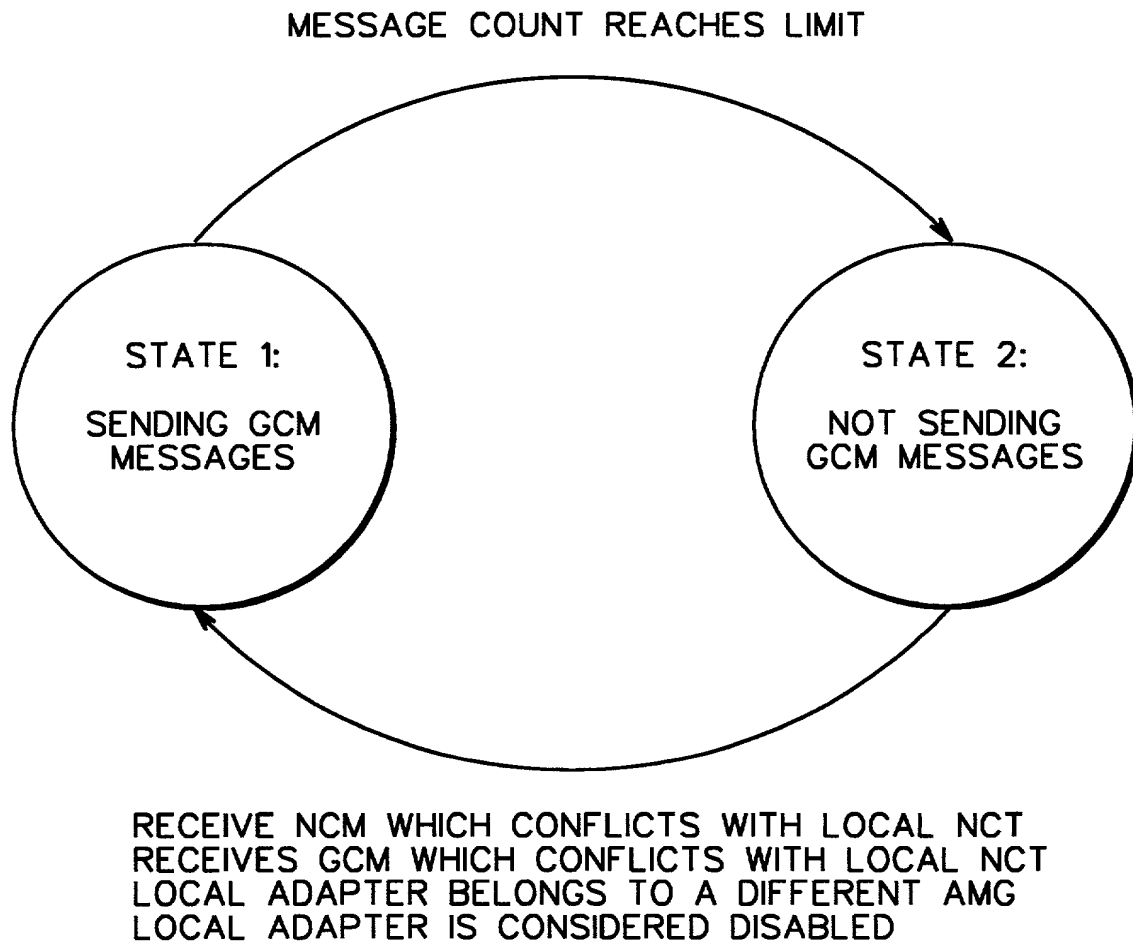


fig. 7

Docket No.
POU9-2000-0017-US1

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**TOPOLOGY PROPAGATION IN A DISTRIBUTED COMPUTING ENVIRONMENT WITH
NO TOPOLOGY MESSAGE TRAFFIC IN STEADY STATE**

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on _____ as United States Application No. or PCT International
Application Number _____
and was amended on _____

(if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Not Claimed

(Number)

(Country)

(Day/Month/Year Filed)

☐

(Number)

(Country)

(Day/Month/Year Filed)

☐

(Number)

(Country)

(Day/Month/Year Filed)

☐

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. *(list name and registration number)*

Lynn L. Augspurger, Reg. No. 24,227

Lawrence D. Cutter, Reg. No. 28,501

Marc A. Ehrlich, Reg. No. 39,966

William B. Porter, Reg. No. 33,135

Floyd A. Gonzalez, Reg. No. 26,732

William A. Kinnaman, Jr., Reg. No. 27,650

Lily Neff, Reg. No. 38,254

Andrew J. Wojnicki, Jr., Reg. No. 43,995

Christopher A. Hughes, Reg. No. 26,914

Edward A. Pennington, Reg. No. 32,588

John E. Hoel, Reg. No. 26,279

Joseph C. Redmond, Jr., Reg. No. 18,753

Jeff Rothenberg, Reg. No. 26,429

Kevin P. Radigan, Reg. No. 31,789

Blanche E. Schiller, Reg. No. 35,670

Send Correspondence to: **Kevin P. Radigan, Esq.**
HESLIN & ROTHENBERG, P.C.
5 Columbia Circle
Albany, NY 12203

Direct Telephone Calls to: *(name and telephone number)*
Kevin P. Radigan, Esq. - (518)452-5600

Full name of sole or first inventor Felipe Knop	
Sole or first inventor's signature <i>Felipe Knop</i>	Date 05/24/2000
Residence 9 Lafayette Court, Poughkeepsie, New York 12603	
Citizenship Brazil	
Post Office Address 9 Lafayette Court, Poughkeepsie, New York 12603	

Full name of second inventor, if any Dennis D. Jurgensen	
Second inventor's signature	Date
Residence 170 Hasbrouck Drive, Apex, North Carolina 27502	
Citizenship United States of America	
Post Office Address 170 Hasbrouck Drive, Apex, North Carolina 27502	

Full name of third inventor, if any Chun-Shi Chang	
Third inventor's signature <i>Chunsh Chang</i>	Date 05/24/2000
Residence 6 Saddle Rock Drive, Poughkeepsie, New York 12603	
Citizenship Taiwan (a/k/a Republic of China)	
Post Office Address 6 Saddle Rock Drive, Poughkeepsie, New York 12603	

Full name of fourth inventor, if any Sameh A. Fakhouri	
Fourth inventor's signature	Date
Residence 143 Storer Avenue, New Rochelle, New York 10801	
Citizenship United States of America	
Post Office Address 143 Storer Avenue, New Rochelle, New York 10801	

Full name of fifth inventor, if any Timothy L. Race	
Fifth inventor's signature <i>Timothy L. Race</i>	Date 05/24/2000
Residence 304 Prince Lane, Kingston, New York 12401	
Citizenship United States of America	
Post Office Address 304 Prince Lane, Kingston, New York 12401	

Full name of sixth inventor, if any	
Sixth inventor's signature	Date
Residence	
Citizenship	
Post Office Address	

Docket No.
POU9-2000-0017-US1

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

TOPOLOGY PROPAGATION IN A DISTRIBUTED COMPUTING ENVIRONMENT WITH NO TOPOLOGY MESSAGE TRAFFIC IN STEADY STATE

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on _____ as United States Application No. or PCT International
Application Number _____
and was amended on _____
(if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Not Claimed

_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

Lynn L. Augspurger, Reg. No. 24,227

Lawrence D. Cutter, Reg. No. 28,501

Marc A. Ehrlich, Reg. No. 39,966

William B. Porter, Reg. No. 33,135

Floyd A. Gonzalez, Reg. No. 26,732

William A. Kinnaman, Jr., Reg. No. 27,650

Lily Neff, Reg. No. 38,254

Andrew J. Wojnicki, Jr., Reg. No. 43,995

Christopher A. Hughes, Reg. No. 26,914

Edward A. Pennington, Reg. No. 32,588

John E. Hoel, Reg. No. 26,279

Joseph C. Redmond, Jr., Reg. No. 18,753

Jeff Rothenberg, Reg. No. 26,429

Kevin P. Radigan, Reg. No. 31,789

Blanche E. Schiller, Reg. No. 35,670

Send Correspondence to: Kevin P. Radigan, Esq.
HESLIN & ROTHENBERG, P.C.
5 Columbia Circle
Albany, NY 12203

Direct Telephone Calls to: (name and telephone number)
Kevin P. Radigan, Esq. - (518)452-5600

Full name of sole or first inventor Felipe Knop	
Sole or first inventor's signature	Date
Residence 9 Lafayette Court, Poughkeepsie, New York 12603	
Citizenship Brazil	
Post Office Address 9 Lafayette Court, Poughkeepsie, New York 12603	

Full name of second inventor, if any Dennis D. Jurgensen	
Second inventor's signature <i>Dennis D. Jurgensen</i>	Date 5/25/2000
Residence 170 Hasbrouck Drive, Apex, North Carolina 27502	
Citizenship United States of America	
Post Office Address 170 Hasbrouck Drive, Apex, North Carolina 27502	

Docket No.
POU9-2000-0017-US1

Declaration and Power of Attorney For Patent Application

English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

**TOPOLOGY PROPAGATION IN A DISTRIBUTED COMPUTING ENVIRONMENT WITH
NO TOPOLOGY MESSAGE TRAFFIC IN STEADY STATE**

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on _____ as United States Application No. or PCT International
Application Number _____
and was amended on _____
(if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Not Claimed

_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	<input type="checkbox"/>

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

(Application Serial No.)

(Filing Date)

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

(Application Serial No.)

(Filing Date)

(Status)
(patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. (list name and registration number)

Lynn L. Augspurger, Reg. No. 24,227
Lawrence D. Cutter, Reg. No. 28,501
Marc A. Ehrlich, Reg. No. 39,966
William B. Porter, Reg. No. 33,135
Floyd A. Gonzalez, Reg. No. 26,732
William A. Kinnaman, Jr., Reg. No. 27,650
Lily Neff, Reg. No. 38,254
Andrew J. Wojnicki, Jr., Reg. No. 43,995

Christopher A. Hughes, Reg. No. 26,914
Edward A. Pennington, Reg. No. 32,588
John E. Hoel, Reg. No. 26,279
Joseph C. Redmond, Jr., Reg. No. 18,753
Jeff Rothenberg, Reg. No. 26,429
Kevin P. Radigan, Reg. No. 31,789
Blanche E. Schiller, Reg. No. 35,670

Send Correspondence to: Kevin P. Radigan, Esq.
HESLIN & ROTHENBERG, P.C.
5 Columbia Circle
Albany, NY 12203

Direct Telephone Calls to: (name and telephone number)
Kevin P. Radigan, Esq. - (518)452-5600

Full name of sole or first inventor Felipe Knop	
Sole or first inventor's signature	Date
Residence 9 Lafayette Court, Poughkeepsie, New York 12603	
Citizenship Brazil	
Post Office Address 9 Lafayette Court, Poughkeepsie, New York 12603	

Full name of second inventor, if any Dennis D. Jurgensen	
Second inventor's signature	Date
Residence 170 Hasbrouck Drive, Apex, North Carolina 27502	
Citizenship United States of America	
Post Office Address 170 Hasbrouck Drive, Apex, North Carolina 27502	

Full name of third inventor, if any Chun-Shi Chang	
Third inventor's signature	Date
Residence 6 Saddle Rock Drive, Poughkeepsie, New York 12603	
Citizenship Taiwan (a/k/a Republic of China)	
Post Office Address 6 Saddle Rock Drive, Poughkeepsie, New York 12603	

Full name of fourth inventor, if any Samah A. Fakhouri	
Fourth inventor's signature <i>Samah A. Fakhouri</i>	Date 5/24/2000
Residence 143 Storer Avenue, New Rochelle, New York 10801	
Citizenship United States of America	
Post Office Address 143 Storer Avenue, New Rochelle, New York 10801	

Full name of fifth inventor, if any Timothy L. Race	
Fifth inventor's signature	Date
Residence 304 Prince Lane, Kingston, New York 12401	
Citizenship United States of America	
Post Office Address 304 Prince Lane, Kingston, New York 12401	

Full name of sixth inventor, if any	
Sixth inventor's signature	Date
Residence	
Citizenship	
Post Office Address	